

Evolution of the Internet k -dense structure

Chiara Orsini, Enrico Gregori, Luciano Lenzini, Dmitri Krioukov

Abstract—As the Internet AS-level topology grows over time, some of its structural properties remain unchanged. Such time-invariant properties are generally interesting, because they tend to reflect some fundamental processes or constraints behind Internet growth. As has been shown before, the time-invariant structural properties of the Internet include some most basic ones, such as the degree distribution or clustering. Here we add to this time-invariant list a non-trivial property— k -dense decomposition. This property is derived from a recursive form of edge multiplicity, defined as the number of triangles that share a given edge. We show that after proper normalization, the k -dense decomposition of the Internet has remained stable over the last decade, even though the Internet size has approximately doubled, and so has the k -density of its k -densest core. This core consists mostly of content providers peering at Internet eXchange Points, and it only loosely overlaps with the high-degree or high-rank AS core, consisting mostly of tier-1 transit providers. We thus show that high degrees and high k -densities reflect two different Internet-specific properties of ASes (transit versus content providers), thus explaining strong fluctuations between degrees and k -densities, and the related observation that random graphs with the same degree distribution or even degree correlations as in the Internet, do not reproduce its k -dense decomposition. Therefore an interesting open question is what Internet topology models or generators can fully explain or at least reproduce the k -dense properties of the Internet.

Index Terms—Internet topology, network evolution, k -dense, dK -graphs.

I. INTRODUCTION

The discovery of power laws in the Internet in 1999 [1] came as a big surprise to many, and as a source of major disbelief to some. Even more surprising is that over the last decade, an increasing number of increasingly refined and complete macroscopic Internet topology measurements and data sources (with one exception, WHOIS) show that the power-law distribution of AS degrees in the Internet has remained exceptionally stable, i.e. time-invariant [2], [3], [4], [5], [6], [7]. One has to always keep in mind that these macroscopic measurements may miss a significant percentage of links. Indeed, there have been several studies supporting this expectation: huge percentages of links are reported as missing in [8], [9], [10], for example. These studies rely on proprietary data collected from only a few ASes. Therefore questions about statistical significance of the reported results, which are difficult or impossible to reproduce due to the

proprietary nature of the data, are well grounded. However these questions may be not so important in view of that all these studies report that a majority of missing links are peering, while customer-provider links appear to be covered well in macroscopic topology measurements. These results are expected, given an ease at which peering links are “set up” at Internet eXchange Points (IXPs), for example: as soon as an AS connects to an IXP and declares an open peering policy, it can exchange traffic with any other openly peering AS at the same IXP. It is instructive to compare the setup of such “links” with the process behind setting up customer-provider links, often involving complicated business decisions, negotiations, and payment agreements [11], [12].

The time-invariant nature of the power-law distribution of (customer-provider) degrees, as well as of strong clustering, has recently found an interesting explanation in certain trade-off optimization drivers behind Internet evolution [13]. These drivers are truly fundamental and apply not only to the Internet, but also to social and biological networks, and even to spacetime in our accelerating universe [14]. More generally, any time-invariant property of an evolving network structure is a candidate to reflect some fundamental forces or constraints behind network evolution, unless this property is a simple statistical consequence of some other property—degree distribution, for example. These forces and constraints can be quite general, applicable to many different networks, as is the case with degree distribution and clustering, or they can be specific to a given network, in which case they likely reflect some specific functions that this network performs.

Here we show that the Internet k -dense decomposition is time-invariant, and explain this invariance via Internet-specific AS data reflecting different functions and business roles of different ASes. The k -dense decomposition [15] of a graph G is a hierarchy of nested subgraphs H_k , $H_k \subset H_{k-1} \subset G$, $k = 2, 3, \dots, k_{MAX}$, induced by edges belonging to $k - 2$ or more triangles within H_k . If a link belongs to H_k but not to H_{k+1} , it is said to have the k -dense-index equal to k . The k -dense-index of a node is the maximum k -dense-index of its incident edges. All other definitions are in Section III. The k -dense-index of an edge is thus a recursive variant of edge multiplicity, defined as the number of triangles in G that an edge is a member of. Edge multiplicity was introduced and studied in [16], [17], where it was also shown that the edge multiplicity distribution in many real networks is either power-law or fat-tailed, and that many existing network models fail to reproduce this property. The k -dense decomposition of one-time Internet snapshots have been analyzed in [18]. Here we collect historical Internet topology data from May 2004 to May 2012, Section II, to study Internet evolution from the k -dense decomposition perspective. Our main results include:

- 1) The maximum k -dense-index k_{MAX} exhibits a clear

C. Orsini is the corresponding author. She is with CAIDA/UCSD, IIT/CNR and IET/UNIPI. e-mail: chiara@caida.org

E. Gregori is with the Institute of Informatics and Telematics, Italian National Research Council (IIT/CNR), Pisa, Italy. e-mail: enrico.gregori@iit.cnr.it

L. Lenzini is with the Department of Information Engineering, University of Pisa (IET/UNIPI), Pisa, Italy. e-mail: l.lenzini@iet.unipi.it

D. Krioukov is with the Cooperative Association for Internet Data Analysis, University of California San Diego (CAIDA/UCSD), San Diego, CA, USA. e-mail: dima@caida.org

growing trend, increasing from 29 in 2004 to 48 in 2012, Section IV-A, while the number of ASes in the densest core $H_{k_{MAX}}$ is small and fluctuating—it is 59 in 2004 and 60 in 2012. The k_{MAX} growth can only partially be attributed to the growing average degree. Other factors are likely related to increasing open peering at IXPs: all the 60 ASes in the 2012 $H_{k_{MAX}}$ are present at at least one IXP, Section V-A.

- 2) After proper normalization, Section IV-B, the Internet k -dense decomposition is shown to be time-invariant in Section IV-C. The Internet-specific interpretation of this result appears in Sections V-B and V-C.
- 3) A significant part of this interpretation is centered around the observation that the degree of an AS and its k -density reflect two different Internet-specific properties of the AS, two different AS business roles. Related results include:
 - a) While some correlations between AS degree and k -density are present as expected, the fluctuations between the two quantities are very strong, Section IV-D.
 - b) Random graphs having the same degree distribution or even degree correlations as the Internet, do not have the same k -dense decomposition, Section IV-E.
 - c) The analysis of PeeringDB data reveals that while high-degree or high-rank ASes tend to be tier-1 transit providers, high- k -density ASes tend to be tier-2 and content providers peering at IXPs, Section V-C.
- 4) The structure of densest core $H_{k_{MAX}}$ is statistically determined by its degree distribution alone, Section IV-F. We interpret this result via open peering policies of participating ASes, which do not choose their peers based on their degrees, Section V-D. On the contrary, selective peering policies, present elsewhere in the Internet, likely introduce degree correlations, providing a new interpretation of the main result in [19], where the AS-level topology of the Internet was found to be statistically determined by its degree correlations.

Some other **related work** dealing with Internet evolution includes [5], which evaluates how different structural properties, related to the distributions of node degrees, centralities, path lengths, community structure, etc., change over time, from January 2002 to January 2010. The study relies on the Cramér-von Mises criterion to identify changes between the distributions, and finds that most distributions remain unchanged, except for the average path length and clustering coefficient. These changes are interpreted as a consequence of peering policy changes. The different growth dynamics of the IPv4 and IPv6 topologies from 1997 to 2009 are juxtaposed in [20]. The main result is that IPv4 topology growth had a phase transition in 2001, while IPv6 had a different phase transition in 2006. The authors of [21] focus on topology liveness and completeness problems, comparing different Internet topology measurement data sources for the period from January 2004 to December 2006. Two evolution trends are highlighted in this

work: a) customer networks are the major cause of the overall topology growth, b) transit providers tend to form increasingly denser structures. The monumental study [3] analyzes the evolution of customer-provider connections in the Internet from January 1998 to January 2010. AS links and nodes are labeled by business relationships and roles, and studied separately. The authors find that enterprise networks and content/access providers at the periphery are the main contributors to the overall Internet growth. They also study rewiring activity, and find that content/access providers appear as most active in that regard. The evolution of the k -core decomposition of the Internet from December 2001 to December 2006 is studied in [22]. Similar to the k -dense decomposition that we analyze here, the k -core decomposition also appears time-invariant according to [22]. However, contrary to the maximum k -density, the maximum k -coreness does not exhibit any growing trend. To the best of our knowledge, it remains unclear if these results in general, and AS coreness in particular, can find any Internet-specific interpretations.

II. DATA

We use the AS-level topology data from the UCLA Computer Science Department's Internet Research Lab (IRL) [23], [24]. We collect yearly Internet snapshots dated from May 2004 to May 2012. Specifically, for each year, we download the data corresponding to May 31st, and then discard all the links with the *last seen* attribute older than May 1st. Table I reports the numbers of AS nodes and links in each snapshot.

We emphasize that BGP and traceroute-based data provide incomplete and biased views of the real Internet AS-level topology. For example, many connections between leaf (low-degree) ASes are hidden from monitors located in hub (high-degree) ASes [25]. Changing numbers of monitors introduce various artifacts and aberrations in the observed topologies. Since only the best paths are typically announced and used for routing, the observed views are highly incomplete. Specific to traceroute measurements, there are many non-responding ASes, especially leaf ASes comprising a majority of Internet ASes. There are also many issues with mapping IP addresses to AS numbers, including IP addresses for which there are multiple or no mapping ASes, and many other vagaries [26].

Our decision to use the IRL BGP data was motivated primarily by the observation that the number of ASes in this dataset is relatively consistent with the number of the Advertised AS Count growth in the CIDR report [27].

III. k -DENSE DEFINITIONS

The k -dense decomposition is a recursive graph decomposition [15], based on edge multiplicity [16], [17]. The multiplicity $m_G(i, j)$ of edge (i, j) in graph G is the number of triangles in G containing the edge, or equivalently, the number of common neighbors of connected nodes i and j , see Figure 1. By definition, the k -dense subgraph H_k of graph G is the subgraph induced by all the links with multiplicity larger or equal to $k - 2$ in the subgraph:

$$m_{H_k}(i, j) \geq k - 2. \quad (1)$$

TABLE I
NUMBER OF AS NODES AND LINKS.

	2004	2005	2006	2007	2008	2009	2010	2011	2012
Nodes	17,858	20,486	23,044	26,101	29,042	32,379	35,583	38,888	42,419
Links	50,326	59,382	66,781	79,931	90,588	102,362	113,846	127,558	146,271

This subgraph can be obtained from G by iterative pruning all the links with multiplicity smaller than $k - 2$. All the links in H_k have multiplicity equal or greater than $k - 2$ in G as well, $m_{H_k}(i, j) \geq k - 2 \Rightarrow m_G(i, j) \geq k - 2$, but the converse is not generally true. That is, links with $m_G(i, j) \geq k - 2$ are only *candidate* H_k links.

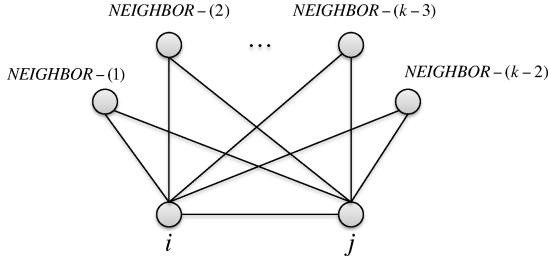


Fig. 1. Link i, j of multiplicity $m(i, j) = k - 2$.

The k -dense decomposition of G is a set of nested subgraphs $H_{k_{MAX}} \subset \dots \subset H_{k+1} \subset H_k \subset \dots \subset G$, where $H_{k_{MAX}}$ is the smallest and densest non-empty sub-graph.

A link is said to have the k -dense-index equal to k^* if it belongs to H_{k^*} but not to H_{k^*+1} . The set of all the links with the k -dense-index equal to k^* is called the k^* -dense-shell.

A node is said to have the k -dense-index equal to k^* if k^* is the maximum k -dense-index of its incident links. The set of all the nodes with k -dense-index equal to k^* is called the k^* -dense-set.

Simply put, the k -dense decomposition shows how densely connected the nodes are, and how nodes belonging to different k -dense sets interconnect. Compared to similar k -core [28] or k -clique [29] decompositions: a) the k -dense decomposition suggests a stronger relationships among nodes belonging to the same k -set; b) the k -dense decomposition is less sensitive to noise than the k -clique decomposition; c) from a computational standpoint, the k -dense decomposition is slightly more complex than k -core but much less complex than k -clique.

IV. INTERNET'S k -DENSE DECOMPOSITION

Treating the data described in Section II as a sorted set of AS-level graphs ordered by year, we report in this Section the results of the statistical analysis of Internet's k -dense decomposition and its evolution over time. The interpretation of these results is deferred to Section V.

A. Basic trends

Similar to other studies, e.g. [21], [3], we have not observed and do not report any significant changes in the basic graph properties—degree distribution, degree correlations, clustering, betweenness, and shortest path distributions—even though

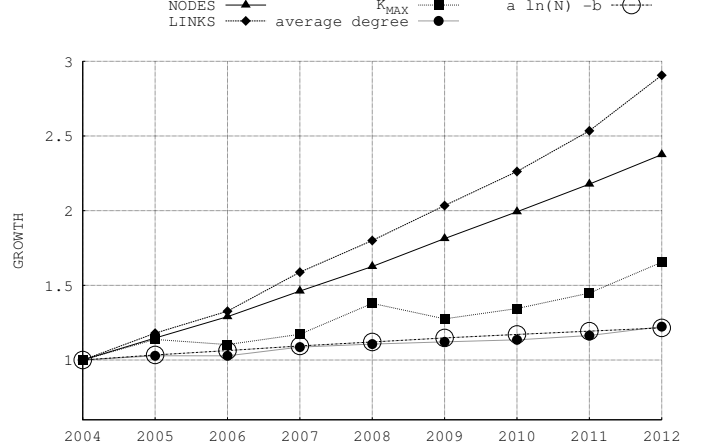


Fig. 2. Internet growth in terms of numbers of nodes, links, k_{MAX} -dense-index and average degree. The black triangles show $N(t)/N(t_0)$, the number of nodes in the graph at time t divided by the number of nodes at time $t_0 = 2004$, $N(t_0) = 17,858$. The black rhombuses show the same ratio for the number of links, $M(t)/M(t_0)$, $M(t_0) = 50,326$. The black squares are $k_{MAX}(t)/k_{MAX}(t_0)$, $k_{MAX}(t_0) = 29$. The black circles show the average degree of the graph at time t divided by the average degree at time $t_0 = 2004$, $\bar{k}(t_0) = 5.64$. The average degree grows logarithmically with the Internet size, $\bar{k} \approx a \ln N - b$, with $a = 1.3$ and $b = 7.5$. The empty circles are $(a \ln N(t) - b)/(a \ln N(t_0) - b)$, showing the quality of this approximation.

the graph has grown significantly, Table I and Figure 2. We see that the number of links M has been growing faster than the number of nodes N , meaning that the average degree $\bar{k} = 2M/N$ has been increasing. This increase appears to be a logarithmic function of N , $\bar{k} \approx a \ln N - b$, with $a = 1.3$ and $b = 7.5$, in agreement with [13], [14]. The increasing trend of k_{MAX} growing from 29 in 2004 to 48 in 2012 indicates that more densely connected parts in the Internet have been forming in the course of its growth.

B. k -dense normalization

For each snapshot we first compute the number of nodes and links with a given k -dense-index, i.e. the number of nodes in each k -dense-set and the number of links in each k -dense-shell. Since the graph in the snapshots and its k_{MAX} are growing, to be able to properly compare the nine snapshots we next perform the following normalization, mapping k -dense-indices and corresponding numbers of nodes and links to fractions with values between 0 and 1:

- *x-axis normalization*: map each index k in each snapshot to what we call the *k-dense-index fraction*:

$$x = \frac{k - k_{MIN}}{k_{MAX} - k_{MIN}}, \quad (2)$$

where k_{MIN} and k_{MAX} are the minimum and maximum values of the k -dense-index in the snapshot;

- *y-axis normalization*: divide the corresponding number of nodes or links by the total number of nodes or links in the snapshot.

In all the considered snapshots there are no nodes of degree 0, so that $k_{MIN} = 2$ for links and nodes in all the snapshots. This means that $x = 0$ corresponds to the k -dense-index equal to 2, while $x = 1$ corresponds to the k -dense-index equal to k_{MAX} , which grows with time. Upon such normalization it becomes difficult to read off the absolute values of k -dense-indices, nodes, or links from the resulting plots, but this normalization is helpful to see if there are any size- and time-invariant statistical trends.

C. k -dense decomposition

In Figures 3a and 3b we report the normalized fractions of nodes and links for each k -dense-index, averaged over the nine historical snapshots. These average values are highly representative for each snapshot from 2004 to 2012, in the sense that the distributions of the actual values across the snapshots are quite narrow, as indicated by the 80% percentile and min/max deviations from the averages in the figures. In other words, the shown normalized statistics are fairly stable over time, i.e. time- and size-invariant. In particular, a vast majority of nodes and links in all the snapshots have low k -dense-indices, not belonging to any dense communities.

Similar to the joint degree distribution, showing how nodes of different degree interconnect [19], we next focus on how different k -dense-sets interconnect. In Figure 3c we report the normalized fraction of links attached to nodes in a given k -dense-set. Similarly to Figures 3a and 3b, we observe that this statistics is also stable over time. We also observe that there are two classes of k -dense-sets to which a vast majority of all links are attached—those with the smallest and largest k -dense-index fractions close to 0 and 1. Upon examination of the original data we find that the left peak in Figure 3c is formed by the links attached to the 2- and 3-dense sets, while the right peak is due to the links attached to the k_{MAX} -dense set. This observation motivates us to further focus on the links attached to these three sets, and show where their other ends go in Figures 3d-3f.

We see that 2- and 3-dense-attached links go mostly to other low-dense-index nodes and to the nodes in the k_{MAX} -dense-set, and that k_{MAX} -dense nodes direct a considerable percentage of their connections not only to nodes in low-dense-sets, but also to nodes in densely-connected parts of the network. We also notice that all the considered normalized statistics appear to be quite stable over time, as indicated by the narrow 80% percentiles.

Given that a vast majority of links are attached to nodes in 2-, 3-, or k_{MAX} -dense sets, we further characterize these sets in Table II. The significant percentages of links attached to 2- and 3-dense sets are not surprising in view of that a majority of nodes belong to these sets, Figure 3a. We see in Table II that these peripheral sets are populated with nodes of low average degree and betweenness centrality of the order of

TABLE II
AVERAGE PROPERTIES OF NODES IN THE 2-, 3-, AND k_{MAX} -DENSE-SETS:
AVERAGE DEGREE \bar{k} , AVERAGE CLUSTERING COEFFICIENT \bar{c} , AND
AVERAGE BETWEENNESS CENTRALITY \bar{b} .

	\bar{k}	\bar{c}	\bar{b}
2-dense-set	1.643	0	$2.1 \cdot 10^3$
3-dense-set	3.161	0.746	$9.3 \cdot 10^3$
k_{MAX} -dense-set	403.8	0.194	$3.0 \cdot 10^6$

the number of nodes N in the graph. The k_{MAX} -dense-set, on the other hand, consists of a small number of nodes with high average degree and betweenness of the order of N^2 . That is, these nodes have a central position in the network, they are a key element shaping the overall connectivity, with many links attached to them, Figure 3c, which motivates us to further analyze the structure of this k_{MAX} -dense subgraph. However before doing so in Section IV-F, we first answer the following two natural questions:

- 1) Does the degree of a node fully define its k -dense-index, making this index a repetitive statistics providing no new information compared to node degree?
- 2) Do random graphs having the same degree distribution or degree correlations as the Internet, fully reproduce all its k -dense properties, making them statistically insignificant, that is, casting them as statistical consequences of that this graph has the observed (joint) degree distribution?

D. k -dense-index versus node degree

If all nodes of the same degree have the same k -index, then one determines the other, so that the second statistics has no independent value. In Figure 4 we show the relationship between k -dense-indices of nodes and their degrees in the 2012 snapshot. As expected the average degree of nodes in a given k -index-set as a function of k exhibits a growing trend, however the fluctuations are high. In particular, the degrees of nodes in the 2-dense-set vary from 2 to 45, while the degree of nodes in the k_{MAX} -dense-set ($k_{MAX} = 48$) vary from 103 to 2964. The other way around, nodes of degree 100, for example, have k -indices ranging from 12 to 45, and the node with maximum degree 174 (Cogent Communications inc., $k = 3655$) has the k -dense-index equal to $32 < k_{MAX}$. We conclude that the k -index is not fully determined by the node degree, an observation that gains further support and Internet-specific explanation in Section V.

E. Statistical significance of the k -dense properties

Even though the node degree does not fully determine the node k -index, it may still be the case that random graphs having the same degree distribution as the Internet, also have exactly the same k -dense properties, meaning that these properties are nothing but statistical artifacts, as it turned out to be the case with the rich club connectivity in the Internet [30]. To check if a similar story applies to the k -dense properties, we construct dK -random graphs [19] for $d = 0, 1, 2$ as described in Section IV-F. We generate 10 realizations for each d . These

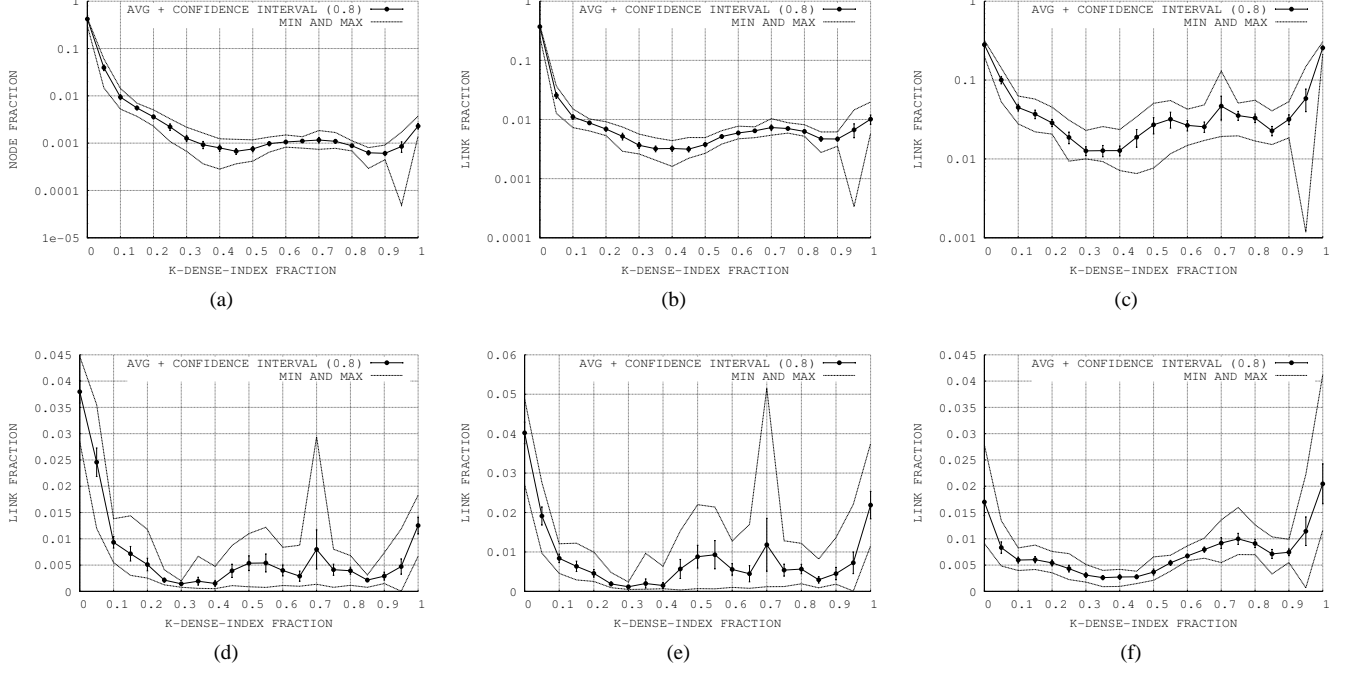


Fig. 3. Normalized k -dense decomposition of the Internet. The x -axes are linearly binned with the bin size equal to 0.05, and each plot shows the average (black dots and solid lines), 80% percentile (vertical bars), and minima and maxima (dashed lines) of the corresponding values computed across the nine historical snapshots of the Internet. (a) is the fraction of nodes with a given k -dense-index, i.e. the number of nodes in the k -dense-set divided by the total number of nodes in the corresponding snapshot. (b) is the fraction of links with a given k -dense-index, i.e. the number of links in the k -dense-shell divided by the total number of links in the corresponding snapshot. (c) is the fraction of links attached to a given k -dense-set, i.e. the number of links whose one or two ends are attached to nodes in a given k -dense-set, divided by the total number of links in the corresponding snapshot. (d) is the number of links with one end attached to a 2-dense-set node and with the other end attached to a k -dense-set node, divided by the number of links attached to the 2-dense-set. (e) and (f) show the corresponding fractions for links attached to 3- and k_{MAX} -dense-sets.

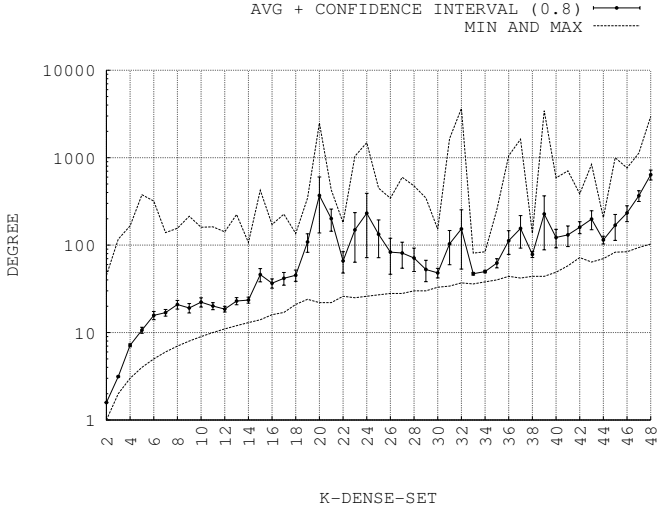


Fig. 4. Distributions of node degrees in k -dense-sets in the 2012 Internet snapshot. The figure shows the averages, 80% percentiles, and min/max values for each of the 48 distributions.

graphs are random graphs with the same average degree, degree distribution, or joint degree distribution as in the Internet 2012 snapshot. In Table III we report the distributions of the k_{MAX} values in these random graphs for each d , while in Figure 5 we show the k -dense link decompositions of these graphs, and juxtapose them against the Internet's. We observe

TABLE III

k_{MAX} -DENSE-INDEX IN INTERNET 2012 AND ITS dK -RANDOMIZATIONS FOR $d = 0, 1, 2$. THE TABLE SHOWS THE AVERAGES \bar{k}_{MAX} AND STANDARD DEVIATIONS σ OF THE DISTRIBUTIONS OF THE k_{MAX} -INDEX VALUES IN dK -RANDOM GRAPH INSTANCES FOR $d = 0, 1, 2$.

	\bar{k}_{MAX}	σ
Internet	48	-
0K-random	3	0
1K-random	68	4.05
2K-random	44.3	0.46

that neither degree distribution nor joint degree distribution fully reproduce the k -dense properties of the Internet, meaning that these properties have their own statistical significance, which is another observation that finds a natural Internet-specific explanation in Section V. Yet another property that finds an explanation rooted in Internet specifics in that section is an unexpectedly simple structure of the k_{MAX} -index core that we analyze next.

F. Structure of the k_{MAX} -dense core

Given the importance of $H_{k_{MAX}}$, the densest and innermost subgraph core, Section IV-C, we next characterize its structure in more detail. In Table IV we show the basic properties of $H_{k_{MAX}}$ subgraphs in all the nine snapshots. We see that these subgraphs are quite small and dense (link density $D = 1$ in complete graphs).

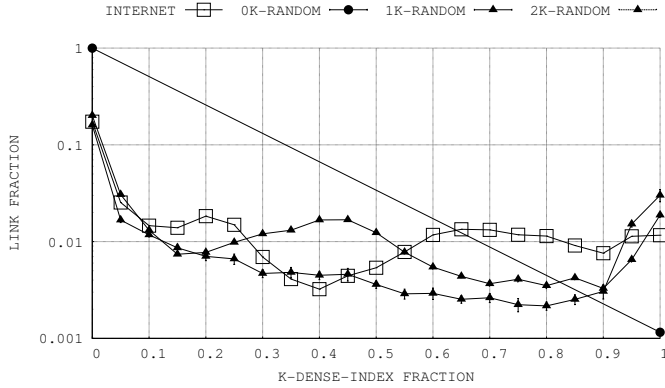


Fig. 5. Normalized k -dense decomposition of links in Internet 2012 and its dK -randomizations for $d = 0, 1, 2$. The figure is similar to Figure 3b, and shows the averages (data points and solid lines) and 80% percentiles (vertical bars) of the distributions of k -dense link fractions in dK -random graph instances for $d = 0, 1, 2$.

TABLE IV

BASIC PROPERTIES OF k_{MAX} -DENSE CORES IN THE NINE HISTORICAL SNAPSHOTS: N THE NUMBER OF NODES; M THE NUMBER OF LINKS, $D = 2M/N/(N-1) \approx k/N$ THE LINK DENSITY.

Year	k_{MAX}	N	M	D
2004	29	59	1,381	0.807
2005	33	55	1,318	0.888
2006	32	44	872	0.922
2007	34	97	3,159	0.678
2008	40	63	1,763	0.903
2009	37	55	1,353	0.911
2010	39	60	1,595	0.901
2011	42	81	2,745	0.847
2012	48	60	1,703	0.962

Similar to the previous section, we next perform the standard dK -statistical analysis [19] of these subgraphs. The procedure is as follows. First we focus on the latest 2012 snapshot, extract the $H_{k_{MAX}}$ subgraph from it, and treat this subgraph as a separate graph. In particular, node degrees are computed within $H_{k_{MAX}}$. Then we construct 20 random graphs having exactly the same average degree as this graph, using the standard $\mathcal{G}_{N,M}$ (Erdős-Rényi [31]) construction procedure of throwing M edges onto $N(N-1)/2$ node pairs uniformly at random. These graphs are called $0K$ -random graphs. Then we construct 20 random graphs having exactly the same degree sequence as $H_{k_{MAX}}$ using the fast generalized Havel-Hakimi algorithm from [32]. This algorithm is guaranteed to always quickly succeed as soon as the degree sequence is graphical, i.e. realizable. The resulting graphs are called $1K$ -random graphs. Then we compute the basic structural graph properties that in neither $0K$ - nor $1K$ -random graphs are guaranteed to be the same as in the original $H_{k_{MAX}}$ graph. The results are in Figure 6.

The average neighbor degree is not guaranteed to be captured by $1K$ -random graphs. Only $2K$ -random graphs, having exactly the same joint degree distribution as the original graph, guarantee to reproduce this degree correlation statistics. Yet we find that $1K$ -random graphs have the same average neighbor degree as $H_{k_{MAX}}$. Clustering is guaranteed to be fully captured only by $3K$ -random graphs, reproducing the frequencies

of triangular subgraphs. Yet we find that $1K$ -random graphs have the same clustering as $H_{k_{MAX}}$. The frequencies of motifs of size 3 or 4 are guaranteed to be fully captured only by $3K$ - or $4K$ -random graphs, the latter reproducing the frequencies of subgraphs of size 4 by definition, but we find that there are no statistically significant deviations of the frequencies of these motifs in $1K$ -random graphs from the corresponding counts in $H_{k_{MAX}}$. Finally the global betweenness or shortest path length distributions are not guaranteed to be captured by dK -random graphs with any $d < N$, but again we see that these global statistics are well reproduced in $1K$ -random graphs, although not as well as the local statistics. At the same time none of the considered structural properties is closely approximated by $0K$ -random graphs (except for the shortest path length distribution).

Collectively these observations imply that even though the $H_{k_{MAX}}$ core is so dense (link density $D = 0.962$) and close to being a complete graph that one could expect it to be well approximated by $0K$ -random graphs, this expectation is actually wrong, but $1K$ -random graphs capture closely the basic structural properties of this core. We applied the same dK -analysis to the $H_{k_{MAX}}$ cores in all the other snapshots, and found them all to be $1K$ -random as well. In other words, $1K$ -randomness of the $H_{k_{MAX}}$ core is another time-invariant property.

V. INTERPRETATION OF INTERNET'S k -DENSE PROPERTIES

In this section we interpret and explain the statistical results from the previous section using specifics of the Internet, where nodes are ASes and links are business relationships between ASes. That is, a link between a couple of ASes is a business agreement between the two organizations that enables these networks to exchange traffic. Recall that business relationships can be very roughly split into two classes: *provider-customer* and *peering*. Providers announce all the destinations to their customers, and thus forward all the traffic that their customers forward to them. Peers mutually announce a limited set of destinations, typically just their own destinations and their customer networks. Providers often charge their customers using the 95th percentile measurement schema [12], i.e. the cost of the service depends on the amount of traffic exchanged. Peering is usually free of charge (unless maintenance costs are considered). The setup of public peering is greatly simplified by Internet eXchange Points (IXPs), facilities where each participant AS can create a single peering connection to any other participants that accept to peer (*open peering* policy), or to a specific subset of those (*selective/restrictive peering* policy). In what follows we explain the main observations from Section IV in view of these Internet-specific realities.

A. Growth of the k_{MAX} -dense index

Even though the size of the $H_{k_{MAX}}$ core fluctuates over time, Table IV, the k_{MAX} -dense index steadily grows, Figure 2. In this section we provide some evidence that this growth is correlated with (if not due to) the proliferation of IXPs.

The first piece of evidence is the structural change of the Internet topology in the last decade due to different growth

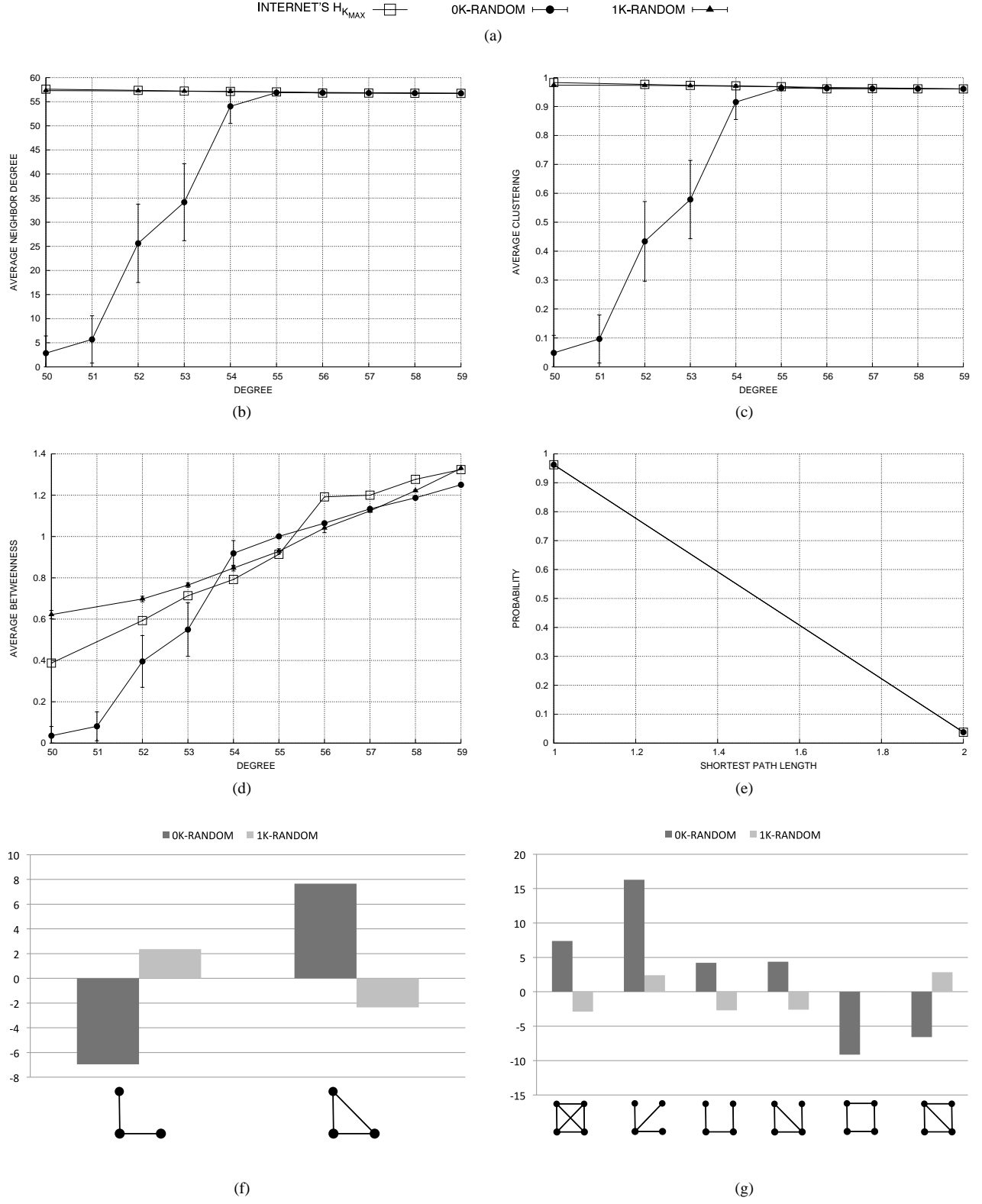


Fig. 6. Basic structural properties of the $H_{k_{MAX}}$ core in Internet 2012 vs. its dK -randomizations for $d = 0, 1$. The figure shows the averages (points and solid lines) and 80% percentiles computed across different dK -random graph instances. Plots (b), (c), and (d) show the average neighbor degree, clustering, and betweenness of nodes of a given degree, while (e) shows the shortest path length distribution. Plots (f) and (g) show the z -scores z for motifs of size 3 and 4, $z = (x - \mu)/\sigma$, where x is a number of occurrences of a given subgraph in $H_{k_{MAX}}$, while μ and σ are the mean and standard deviation of the distribution of the corresponding occurrences computed across different dK -random graph instances.

rates of peering and provider-customer relationships. The original economic driver behind peering is to reduce costs that customers pay to their providers. However, as the price of transit has been steadily decreasing, this driver becomes less prominent, unless large volumes of traffic are exchanged. As shown in [11], peering grows fast, contributing much more to middle AS tiers, compared to the tier-1 ASes. Large Content Providers (CP) and Content Delivery Networks (CDN), which generate a significant percentage of the total Internet traffic [11], [33] are primary drivers behind this process because:

- 1) a shorter path between these networks and subscribers provides better performances;
- 2) although the traffic exchanged is highly asymmetric, for most ISPs the connection to the content providers is vital.

Furthermore, [33] asserts that a dominant percentage of AS-level traffic flows directly between large CPs, CDNs, data centers, and consumer networks, and that 150 ASes originate more than 50% of the Internet inter-AS traffic. Reference [10] also supports our claim by describing the ground truth behind a large European IXP: the authors shows that the amount of peering links established in this facility is unexpectedly huge—about 50,000 peering links among approximately 350 AS members. Yet another piece of evidence can be found in [18]: the percentage of ASes that are members of at least one IXP within a given k -dense set is a rapidly growing function of k . In addition, all the k_{MAX} -dense ASes are members of at least one IXP.

To further support our claim here we perform analysis similar to [18]. We collect the May 2012 information about the 60 k_{MAX} -dense ASes in our 2012 snapshot from PeeringDB [34], a project maintaining a database aimed at facilitating the exchange of information related to peering: “*what networks are peering, where they are peering, and if they are likely to peer with you*”. We find that 58 out of the 60 ASes have a peering record in the database, and that *all* these 58 ASes are members of the Deutscher Commercial Internet Exchange (DE-CIX) [35], one of the largest Internet Exchange Points worldwide, located in Frankfurt, Germany, with the current membership count of more than 450 ASes. We investigate the profile of the remaining 2 ASes, and find that they both are also DE-CIX members: one is a telecommunication company with an unknown peering policy, the other is a hosting service provider with an open peering policy.

B. The 2- and 3-dense-sets

Nodes with the k -dense-index equal to 2 or 3 are a vast majority of all ASes in the Internet. These peripheral ASes contribute most to the overall network growth: they have small degrees, but a large percentage of all connections in the Internet are attached to these ASes [21], [3]. All ASes of degree 1 and all ASes with zero clustering belong to the 2-dense-set. These customer ASes connect to the Internet, but their business is not Internet-driven: they set up their connections to obtain Internet connectivity, thus all they need is a transit provider. Sometimes, for backup, higher availability, or other purposes, they set up agreements with multiple

provider, i.e. they purchase transit from more than one provider (multihoming). If the two providers of a multihomed AS happen to be connected, then a triangle is formed, so that the multihomed AS belongs to the 3-dense-set.

C. The k_{MAX} -dense-set vs. high-rank ASes

The k_{MAX} -dense ASes form the densest-connected community by definition. The easiest way to establish such dense connectivity in practice is by connecting to a large IXP and declaring an open peering policy. Given that CPs and CDNs benefit from peering with any willing-to-peer ASes [37], it is quite plausible that CPs and CDNs are main players behind the formation of this densely interconnected substructure. Surprisingly, the adoption of an open peering policy is an emerging phenomenon among Network Service Providers (NSPs) [38], i.e. tier-2 ASes provided with an own backbone network that purchase transit from an upstream provider and resell it to other ASes. Although these ASes usually adopt a selective peering policy, as they do not want to peer with potential customers, such peering connections help tier-2 ASes to provide a better end-user experience to their customers [39].

The PeeringDB data shown in Tables V(a) and V(d) confirm this statement. Indeed, large percentages of the k_{MAX} -dense ASes:

- can be considered as Content or Network Service Providers,
- direct most of their traffic outbound, and
- have an open peering policy.

The percentage of ASes with selective peering policies (almost all NSPs) is also significant, but all these ASes are good candidates for selective peering as well, explaining their high k -density. Indeed, one commonly considered aspect in the peer selection process is the symmetry of the exchanged traffic [40]. We do not have access to traffic statistics, but we can use the number of IP addresses in the customer cone [41] as a proxy. The customer cone of an AS is the set of ASes that can be reached from the AS following only provider-to-customer links. In other words, it is the set of destinations that can be reached *for free* upon peering with the AS. The distributions of the customer cone sizes (the numbers of /32 IP addresses in the customer cone, to be precise) shown in Figure 7 indicate that the customer cone sizes of k_{MAX} -dense ASes are large, but their distribution is narrower than in the rest of the Internet, thus making k_{MAX} -dense ASes potential peer candidates even in the selective peering case.

We next juxtapose this k_{MAX} -dense-set against different sets of high-rank ASes, in particular high-degree ASes, shedding more light on the different Internet-specific meanings of the k -dense-index and node degree.

Tier-1 is defined as a set of ASes that do not need to pay any transit providers to reach any destination. To accomplish this, all the tier-1 ASes connect to each other forming a clique. This structure ensures that all these ASes have high, although not necessarily *highest* k -dense-indices. However, because they provide routes to all the destinations in the Internet without paying any upstream providers, they have *largest* customer

TABLE V

AS PROPERTIES FROM PEERINGDB. THE k_{MAX} -dense COLUMN REFERS TO THE 60 k_{MAX} -dense ASes IN THE 2012 SNAPSHOT, WHILE THE *degree-rank* AND *AS-rank* COLUMNS REFER TO THE 60 HIGHEST-DEGREE AND HIGHEST-AS-RANK ASes [36]. THE DATA FOR THE HIGHEST-ADDRESS- AND PREFIX-RANK ASes ARE SIMILAR TO THE HIGHEST-AS-RANK ASes.

	k_{MAX} -dense	degree-rank	AS-rank
Cable/DSL/ISP	12	9	5
Content	4	1	0
Educational/Research	1	3	1
Network Service Provider	42	42	43
Non-Profit	1	0	0

(a) Business type.

	k_{MAX} -dense	degree-rank	AS-rank
Asia Pacific	0	0	2
Europe	28	20	8
North America	0	2	5
Regional	17	8	4
Global	15	25	30

(b) Geographic type.

	k_{MAX} -dense	degree-rank	AS-rank
100 Mbps - 10 Gbps	10	6	0
10 Gbps - 1 Tbps	41	36	29
1Tbps+	2	7	12
Unknown	7	11	19

(c) Traffic volume.

	k_{MAX} -dense	degree-rank	AS-rank
Balanced	30	42	37
Mostly Inbound	7	6	9
Mostly Outbound	21	7	3
Unknown	2	5	11

(d) Traffic ratio.

	k_{MAX} -dense	degree-rank	AS-rank
Open	32	15	7
Restrictive	1	11	15
Selective	26	29	27
Unknown	1	5	11

(e) Peering policy.

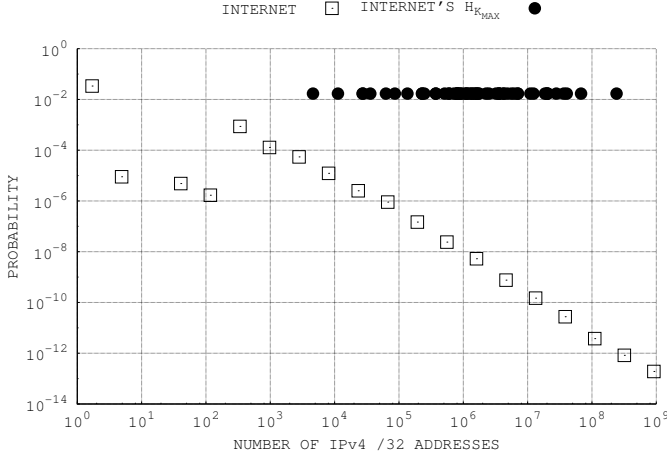


Fig. 7. Distributions of customer cone sizes for the k_{MAX} -dense ASes (black circles) and all ASes in the Internet (empty squares).

cones [41], [8]. Since the main business role of tier-1 ASes (or more generally, of any transit provider) is to sell transit, they have a lot of connections to enterprise customers, Figures 3d and 3e, but the open peering policy would not increase their revenue. Therefore, the tier-1 ASes are more likely to adopt restrictive peering policies. These considerations suggest that these ASes, contrary to a common belief, do not largely belong to the *densest* community, i.e. to $H_{k_{MAX}}$. In particular, the ASes with the largest customer cones are not k_{MAX} -dense, Figure 7.

In addition, we compute the overlaps between the set of the 60 ASes in the $H_{k_{MAX}}$ core of the 2012 snapshot, and the top-60 ASes ranked by their customer cone size. The results in Table VI show that these overlaps are not substantial. In the same table we also show the overlap between the k_{MAX} -max dense ASes and the top-60 highest-degree ASes. From Figure 4 we know that a high degree does not necessarily mean a high k -dense-index. In Table VI we find that the overlap between these two AS sets is about 50%, and if we look back at Table V), we observe that these highest-degree or highest-rank ASes have quite different properties, compared to the k_{MAX} -dense ASes. Specifically, the former AS sets have

- a higher percentage of global ASes, Table V(b),

TABLE VI

THE 60 k_{MAX} -dense ASes IN THE 2012 SNAPSHOT VS. TOP-60 HIGH-RANK ASes. THE TABLE SHOWS THE OVERLAPS BETWEEN THE SET OF ASes IN $H_{k_{MAX}}$ AND: *address-rank*, *prefix-rank*, AND *AS-rank*: THE SETS OF 60 ASes WITH THE LARGEST NUMBERS OF IPv4 / 32 ADDRESSES, IPv4 ROUTING PREFIXES, AND ASes IN THEIR CUSTOMER CONES; AND *degree-rank*: THE SET OF 60 HIGHEST-DEGREE ASes. ALL THE DATA ARE FROM [36].

k -max dense	address-rank	prefix-rank	AS-rank	degree-rank
60	4	6	7	31

- a higher percentage of 10Gbps+ ASes, Table V(c),
- a lower percentage of mostly outbound ASes, Table V(d),
- a higher percentage of ASes having a restrictive peering policy, Table V(e), and
- a lower percentage of ASes having an open peering policy, Table V(e).

Simply put, high-degree or high-rank ASes tend to be very large transit providers, while k_{MAX} -dense ASes tend to be either content providers or tier-2s.

D. 1K-randomness of the $H_{k_{MAX}}$ core

If an AS has an *open peering policy*, it simply sets up a certain number of connections (equal to its degree), without *choosing* its neighbors in any way, e.g. based on their degrees or any other properties. On the contrary, an AS with a *selective/restrictive peering policy* always *chooses* its peers and one of the topological property emerging from this selection process is the degree of the peering candidate, correlated with its customer cone size. In the latter case, we thus cannot expect that the degree distribution alone is sufficient to fully describe the graph. Some non-trivial degree correlations must be present in it, and indeed the Internet AS-level graph as a whole was found to be not 1K- but 2K-random in [19]. The large percentage of open-peering k_{MAX} -dense ASes, and the similarity between their customer cone sizes are thus the main factors explaining the 1K-randomness of the $H_{k_{MAX}}$ subgraph.

Let us assume for a moment that a large fraction of k_{MAX} -dense links is present at a single large IXP. We do not have access to the peering matrix data of any large IXPs, but the data reported in [10] suggest that this assumption may very

well be correct, with DE-CIX being one of possible large IXP candidates. If we also assume that the peering policies declared in PeeringDB reflect the peering policies that each AS follows at each IXP where it has presence, then our explanation of $H_{k_{MAX}}$'s $1K$ -randomness is supported by the data in Figure 7 and Table V(e).

VI. CONCLUSIONS

In summary, as the Internet grows over time, the maximum k -dense index grows as well. That is, the densest and innermost Internet core $H_{k_{MAX}}$ becomes increasingly denser. Yet this form of densification can only partially be attributed to the growing average degree. The fluctuations between node degree and k -dense index are strong, and the two statistics reflect two different Internet-specific properties of AS nodes. High-degree and high- k -dense-index ASes tend to be transit and content providers, respectively. The latter form the densest community in the Internet, which only loosely overlaps with the high-rank ASes forming the tier-1 core. The structure of the $H_{k_{MAX}}$ core is relatively simple. Statistically, it is almost fully determined by its degree distribution, a property that can be explained by open peering policies that many content providers and NSPs tend to follow at IXPs. Most importantly, after proper normalization, all the considered k -dense properties of the Internet appear time-invariant. In particular, a vast majority of all AS links in the Internet are attached either to ASs with the k -dense-index equal to 2 or 3, or to the k_{MAX} -dense ASes.

Speaking more generally, the Internet's k -dense properties, derived from a recursive variant of edge multiplicity measuring the frequency and density of triangle overlaps, appear to be a statistically significant and time-invariant structural properties that cannot be fully captured by either degree distribution ($1K$ -distribution) or degree correlations ($2K$ -distribution). The $3K$ -distribution, the distribution of subgraphs of size 3, may very well capture the Internet's k -dense properties. Even though this conjecture is quite plausible, it is not guaranteed to be correct because reproducing the frequency of degree-labeled triangles does not automatically imply that the whole k -dense decomposition hierarchy of nested subgraphs H_k is correctly reproduced as well. Therefore it is interesting to investigate what existing or new Internet topology models and generators are capable of explaining or at least reproducing the k -dense properties of the Internet.

ACKNOWLEDGMENT

The authors would like to thank Amogh Dhamdhere, Alessandro Improta, Matthew Luckie and Luca Sani for sharing their data, and Hyunju Kim and Zoltan Toroczkai for sharing their code.

REFERENCES

- [1] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," *ACM/SIGCOMM Computer Communication Review*, vol. 29, no. 4, pp. 251–262, Aug. 1999.
- [2] G. Siganos, M. Faloutsos, P. Faloutsos, and C. Faloutsos, "Power Laws and the AS-level Internet Topology," *IEEE/ACM Transactions on Networking*, vol. 11, no. 4, pp. 514–524, Aug. 2003.
- [3] A. Dhamdhere and C. Dovrolis, "Twelve Years in the Evolution of the Internet Ecosystem," *IEEE/ACM Transactions on Networking*, vol. 19, no. 5, pp. 1420–1433, Sep. 2011.
- [4] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*. New York, NY, USA: Cambridge University Press, 2004.
- [5] B. Edwards, S. A. Hofmeyr, G. Stelle, and S. Forrest, "Internet Topology over Time," in *Proceedings of the 8th CS UNM Student Conference*, Apr. 2012.
- [6] B. Huffaker, M. Fomenkov, and kc claffy, "Internet Topology Data Comparison," CAIDA, University of California, San Diego, Tech. Rep., May 2012.
- [7] B. Zhang, R. Liu, D. Massey, and L. Zhang, "Collecting the Internet AS-level Topology," *ACM SIGCOMM Computer Communication Review (CCR) special issue on Internet Vital Statistics*, Jan. 2005.
- [8] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, k. claffy, and G. Riley, "AS relationships: inference and validation," *ACM/SIGCOMM Computer Communication Review*, vol. 37, no. 1, pp. 29–40, Jan. 2007.
- [9] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang, "The (in)completeness of the Observed Internet AS-level Structure," *IEEE/ACM Transactions on Networking*, vol. 18, no. 1, pp. 109–122, Feb. 2010.
- [10] B. Ager, N. Chatzis, A. Feldmann, N. Sarraf, S. Uhlig, and W. Willinger, "Anatomy of a large european IXP," in *Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication*, ser. SIGCOMM '12. New York, NY, USA: ACM, 2012, pp. 163–174.
- [11] B. Krogfoss, M. Weldon, and L. Sofman, "Internet Architecture Evolution and the Complex Economies of Content Peering," *Bell Lab. Tech. J.*, vol. 17, no. 1, pp. 163–184, Jun. 2012.
- [12] W. B. Norton. (2012) Internet Service Providers and Peering v3.0. [Online]. Available: <http://drpeering.net/white-papers/Internet-Service-Providers-And-Peering.html>
- [13] F. Papadopoulos, M. Kitsak, M. Serrano, M. Boguñá, and D. Krioukov, "Popularity versus Similarity in Growing Networks," *Nature*, vol. 489, pp. 537–540, Sep. 2012.
- [14] D. Krioukov, M. Kitsak, R. Sinkovits, D. Rideout, D. Meyer, and M. Boguñá, "Network Cosmology," *Nature Scientific Reports*, vol. 2, no. 793, Nov. 2012.
- [15] K. Saito, T. Yamada, and K. Kazama, "Extracting Communities from Complex Networks by the k -Dense Method," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E91-A, no. 11, pp. 3304–3311, 2008.
- [16] M. A. Serrano and M. Boguñá, "Clustering in complex networks. I. General formalism," *Physical Review E*, vol. 74, 2006.
- [17] V. Zlatić, D. Garlaschelli, and G. Caldarelli, "Networks with arbitrary edge multiplicities," *EPL (Europhysics Letters)*, vol. 97, no. 2, pp. 28005–p1–p5, 2012.
- [18] E. Gregori, L. Lenzini, and C. Orsini, "k-dense Communities in the Internet AS-level Topology Graph," *Computer Networks*, no. 0, pp. –, 2012, (In press).
- [19] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, "Systematic topology analysis and generation using degree correlations," *ACM/SIGCOMM Computer Communication Review*, vol. 36, no. 4, pp. 135–146, Aug. 2006.
- [20] G. Zhang, B. Quoitin, and S. Zhou, "Phase changes in the evolution of the IPv4 and IPv6 AS-Level Internet topologies," *Computer Communications*, vol. 34, no. 5, pp. 649–657, Apr. 2011.
- [21] R. V. Oliveira, B. Zhang, and L. Zhang, "Observing the Evolution of Internet AS Topology," in *Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*, ser. SIGCOMM '07. New York, NY, USA: ACM, 2007, pp. 313–324.
- [22] G.-Q. Zhang, G.-Q. Zhang, Q.-F. Yang, S.-Q. Cheng, and T. Zhou, "Evolution of the Internet and its cores," *New Journal of Physics*, vol. 10, no. 12, p. 123027, 2008.
- [23] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang, "Quantifying the Completeness of the Observed Internet AS-level Structure," UCLA, Tech. Rep. 080026, Sep. 2008.
- [24] UCLA Computer Science Department's Internet Research Lab. (2012) Internet Topology Collection. [Online]. Available: <http://irl.cs.ucla.edu/topology/>
- [25] E. Gregori, A. Improta, L. Lenzini, L. Rossi, and L. Sani, "On the Incompleteness of the AS-level Graph: a Novel Methodology for BGP Route Collector Placement," in *Proceedings of IMC 2012*, 2012.

- [26] K. Keys, Y. Hyun, M. Luckie, and k. claffy, "Internet-Scale IPv4 Alias Resolution with MIDAR," *IEEE/ACM Transactions on Networking*, 2012.
- [27] (2012) CIDR report. [Online]. Available: <http://www.cidr-report.org/as2.0/>
- [28] S. B. Seidman, "Network structure and minimum degree," *Social Networks*, vol. 5, pp. 269–287, 1983.
- [29] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, 2005.
- [30] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, "Detecting rich-club ordering in complex networks," *Nature Physics*, vol. 2, pp. 110–115, 2006.
- [31] P. Erdos and A. Renyi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci.*, vol. 5, pp. 17–61, 1960.
- [32] H. Kim, C. I. D. Genio, K. E. Bassler, and Z. Toroczkai, "Constructing and sampling directed graphs with given degree sequences," *New Journal of Physics*, vol. 14, no. 2, p. 023012, 2012.
- [33] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, "Internet inter-domain traffic," *ACM/SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. –, Aug. 2010.
- [34] (2012) PeeringDB. [Online]. Available: <http://www.peeringdb.com/>
- [35] (2012) DE-CIX, the Deutscher Commercial Internet Exchange. [Online]. Available: <http://www.de-cix.net/>
- [36] CAIDA, the Cooperative Association for the Internet Data Analysis. (2012) The CAIDA AS Relationships Dataset, <2012-06-01>. [Online]. Available: <http://www.caida.org/data/active/as-relationships/>
- [37] W. B. Norton. (2011) Content Providers (aka Content Publishers). [Online]. Available: <http://drpeering.net/white-papers/Ecosystems/Content-Providers.html>
- [38] A. Lodhi, A. Dhamdhere, and C. Dovrolis, "Peering strategy adoption by transit providers in the internet: a game theoretic approach?" *ACM/SIGMETRICS Performance Evaluation Review*, vol. 40, no. 2, pp. 38–41, Oct. 2012.
- [39] W. B. Norton, *The Internet Peering Playbook: Connecting to the Core of the Internet*. DrPeering Press, 2011, ch. The Top Five Motivations to Peer.
- [40] ——. (2012) The Folly of Peering Ratios (as a Peering Candidate Discriminator). [Online]. Available: <http://drpeering.net/white-papers/The-Folly-Of-Peering-Ratios.html>
- [41] CAIDA, the Cooperative Association for the Internet Data Analysis. (2012) AS Rank. [Online]. Available: <http://as-rank.caida.org/>